IT & C

ISSN 2821 - 8469, ISSN – L 2821 - 8469, Volumul 2, Numărul 2, Iunie 2023

Inteligența artificială și securitatea cibernetică - Active IA care se pot constitui în amenințări

Nicolae Sfetcu

Sfetcu, Nicolae (2023), Inteligența artificială și securitatea cibernetică - Active IA care se pot constitui în amenințări, *IT & C*, **2**:2, 45-50, DOI: 10.58679/IT73512, https://www.internetmobile.ro/inteligenta-artificiala-si-securitatea-cibernetica-active-ia-care-se-pot-constitui-in-amenintari/

Publicat online: 05.05.2023

Inteligența artificială și securitatea cibernetică - Active IA care se pot constitui în amenințări

Nicolae Sfetcu
nicolae@sfetcu.com

Artificial Intelligence and Cyber Security - AI Assets That Can Be Threats

Abstract

Artificial intelligence has gained traction in recent years, facilitating intelligent and automated decision-making across deployment scenarios and application areas. While undoubtedly beneficial, it can expose individuals and organizations to new and sometimes unpredictable risks, and open new avenues in attack methods and techniques, as well as creating new data protection challenges. In this regard, a critical element in the threat landscape is the identification of asset categories that may be at risk.

Keywords: artificial intelligence, cyber security, assets, taxonomy

Rezumat

Inteligența artificială a câștigat tracțiune în ultimii ani, facilitând luarea deciziilor inteligente și automate în cadrul scenariilor de implementare și a zonelor de aplicare. Deși este, fără îndoială, benefică, poate expune indivizii și organizațiile la riscuri noi și uneori imprevizibile, și poate deschide noi căi în metodele și tehnicile de atac, precum și crearea de noi provocări în materie de protecție a datelor. În acest sens, un element critic în peisajul amenințărilor este identificarea categoriilor de active pentru care pot fi amenințări.

Cuvinte cheie: inteligența artificială, securitatea cibernetică, active, taxonomia

IT & C, Volumul 2, Numărul 2, Iunie 2023, pp. 45-50
ISSN 2821 - 8469, ISSN – L 2821 – 8469, DOI: 10.58679/IT73512
URL: https://www.internetmobile.ro/inteligenta-artificiala-si-securitatea-cibernetica-active-ia-care-se-pot-constitui-in-amenintari/

Securitatea cibernetică

Inteligența artificială (IA) a câștigat tracțiune în ultimii ani, facilitând luarea deciziilor inteligente și automate în cadrul scenariilor de implementare și a zonelor de aplicare. (1, 2) Asistăm la o convergență a diferitelor tehnologii (de exemplu, Internetul obiectelor, robotică, tehnologii cu senzori etc.) și o cantitate și o varietate în creștere de date, precum și la caracteristicile lor noi (de exemplu, date distribuite) pentru a utiliza IA la scară. În contextul securității cibernetice, IA poate fi văzută ca o abordare emergentă și, în consecință, tehnicile IA au fost utilizate pentru a sprijini și automatiza operațiunile relevante, de ex. filtrarea traficului, analiza criminalistică automatizată etc. Deși este, fără îndoială, benefică, nu ar trebui să ocolim faptul că IA și aplicarea acesteia pentru, de exemplu, luarea automată a deciziilor — în special în implementări critice pentru siguranță, cum ar fi vehiculele autonome, producția inteligentă, eHealth etc.— poate expune indivizii și organizațiile la riscuri noi și uneori imprevizibile, și poate deschide noi căi în metodele și tehnicile de atac, precum și crearea de noi provocări în materie de protecție a datelor.

Inteligența artificială influențează din ce în ce mai mult viața de zi cu zi a oamenilor și joacă un rol cheie în transformarea digitală prin capacitățile sale automate de luare a deciziilor. Beneficiile acestei tehnologii emergente sunt semnificative, dar la fel sunt și preocupările. Prin urmare, este necesar să evidențiem rolul securității cibernetice în stabilirea fiabilității și implementării unei IA de încredere.

Când luăm în considerare securitatea în contextul IA, trebuie să fim conștienți de faptul că tehnicile și sistemele IA pot duce la rezultate neașteptate și pot fi modificate pentru a manipula rezultatele așteptate. Acesta este cazul în special atunci când se dezvoltă software IA care se bazează adesea pe modele de cutie neagră complet (2) sau poate fi chiar folosită cu intenții rău intenționate, de ex. IA ca mijloc de a spori criminalitatea cibernetică și de a facilita atacurile adversarilor rău intenționați. Prin urmare, este *esențial să se securizeze IA în sine. În special, este important:*

- *să se înțeleagă ce trebuie securizat (activele care sunt supuse amenințărilor specifice IA și modelelor adverse);*

- *să se înțeleagă modelele aferente de guvernare a datelor (inclusiv proiectarea, evaluarea și protejarea datelor și procesul de instruire a sistemelor IA);*
- *să se gestioneze amenințările într-un ecosistem multipartid într-un mod cuprinzător, utilizând modele și taxonomii comune;*
- *să se dezvolte controale specifice pentru a se asigura că IA în sine este sigură.*

Inteligența artificială și securitatea cibernetică au o relație multidimensională și o serie de interdependențe. Dimensiunile care pot fi identificate includ următoarele trei:

1. **Securitate cibernetică pentru IA**: lipsa de robustețe și vulnerabilitățile modelelor și algoritmilor IA, de ex. Inferența și manipularea modelelor adverse, atacurile împotriva sistemelor ciberfizice alimentate de IA, manipularea datelor utilizate în sistemele IA, exploatarea infrastructurii de calcul utilizate pentru a alimenta funcționalitățile sistemelor IA, otrăvirea datelor, variații de mediu care provoacă variații în natura intrinsecă a datelor (3), seturi de date de antrenament credibile și de încredere, validare/verificare algoritmică (inclusiv integritatea lanțului de aprovizionare cu software), validarea proceselor de instruire și evaluare a performanței, identificarea credibilă și fiabilă a caracteristicilor, protecția datelor/confidențialitatea în contextul sistemelor IA etc.
2. **IA pentru a sprijini securitatea cibernetică**: IA utilizată ca instrument/mijloc pentru a crea securitate cibernetică avansată prin dezvoltarea unor controale de securitate mai eficiente (de exemplu, firewall-uri active, antivirus inteligent, operațiuni automate CTI (cyber threat intelligence), IA fuzzing, criminalistică inteligentă, scanare e-mail, sandboxing adaptiv, analiză automată de malware, apărare cibernetică automatizată etc.) și pentru a facilita eforturile forțelor de ordine și ale altor autorități publice pentru a răspunde mai bine la criminalitatea informatică, inclusiv analiza creșterii exponențiale a Big Data în contextul investigațiilor, ca și utilizarea abuzivă criminală a IA.
3. **Utilizarea rău intenționată a IA**: utilizarea rău intenționată/adversă a IA pentru a crea tipuri mai sofisticate de atacuri, de ex. malware bazat pe inteligență artificială, inginerie socială avansată, exploatare de conturi de rețele sociale false bazate pe inteligență artificială, atacuri DDoS intensificate cu inteligență artificială, modele generative profunde pentru a crea date false, spargerea parolelor susținută de inteligență artificială etc. Această categorie include atât atacuri vizate de inteligență artificială (concentrate pe subminarea sistemelor IA existente pentru a-și modifica capacitățile), precum și atacurile susținute de IA (cele care includ tehnici bazate pe IA care vizează îmbunătățirea eficacității atacurilor tradiționale).

Securitatea cibernetică poate fi unul dintre fundamentele soluțiilor de inteligență artificială de încredere. Va servi drept o rampă de lansare pentru desfășurarea securizată pe scară largă a IA în întreaga Uniune Europeană. Cu toate acestea, va face acest lucru numai atunci când înțelegerea comună a peisajului amenințărilor relevante și a provocărilor asociate sunt mapate într-o manieră consecventă. Peisajul amenințărilor IA este vast și dinamic, deoarece evoluează alături de inovațiile observate în domeniul IA și de integrarea continuă a numeroase alte tehnologii.

Active IA

Un element critic în peisajul amenințărilor este identificarea categoriilor de active pentru care pot fi amenințări. (1,3) Activele sunt definite ca orice lucru care are valoare pentru o persoană sau organizație și, prin urmare, necesită protecție. În cazul IA, activele sunt, de asemenea, cele care sunt cruciale pentru a răspunde nevoilor pentru care sunt utilizate.

Pe lângă activele generice legate de TIC, cum ar fi datele, software-ul, hardware-ul, rețelele de comunicații, printre altele, IA implică un set de active specifice, cum ar fi modele, procesoare și artefacte care pot fi compromise și/sau deteriorate, fie din cauze intenționate, cât și din cauza lipsei. - cauze intenționate.

Taxonomia activelor

Pentru fiecare dintre etapele ciclului de viață IA, au fost identificate cele mai relevante active, pe baza descrierii funcționale a etapelor specifice și pentru a reflecta componentele IA, dar și activele care sprijină dezvoltarea și implementarea sistemelor IA. Activele includ și procese legate de IA, având în vedere natura lor transversală. Activele au fost clasificate în următoarele 6 categorii (vezi figura de mai jos):

- Date
- Model
- Actori
- Procese
- Mediu/Instrumente
- Artefacte

Tabelul de mai jos ilustrează taxonomia detaliată a activelor pentru IA bazată pe modelul de referință generic al ciclului de viață IA.

PROCESE	MEDIU/INSTRUMENTE	ARTEFACTE
• Ingestia datelor • Stocarea datelor • Explorarea/preprocesarea datelor • Înțelegerea datelor • Etichetarea datelor • Augmentarea datelor • Colectarea datelor • Selectarea caracteristicilor	• Rețele de comunicare • Protocoale de comunicare • Cloud • Platforme ingestia datelor • Platforme explorarea datelor • Instrumente explorarea datelor • DBMS	• Liste de control acces • Utilizare • Propoziții de valoare și modele de afaceri • Cerințe IA informale/semi-formale, model GQM (Scop/Întrebări/Metrici) • Politici de guvernare a datelor

• Selectarea/construcția, instruirea și testarea modelelor • Reglarea modelelor • Adaptarea-învățarea transferului modelelor/Implementarea modelelor • Întreținerea modelelor	• Sisteme de fișiere distribuite • Platforme computaționale • Mediu de dezvoltare integrat • Biblioteci (cu algoritmi pentru dezvoltare, eticjetare, etc.) • Instrumente de monitorizare • Sistem/software de operare • Tehnici de optimizare • Platforme de învățare automată • Procesoare • Instrumente de vizualizare	• Afișarea și reprezentarea grafică a datelor • Parametri statistici descriptivi • Cadru general model, încadrare software, firmware sau hardware • Artefacte compoziție: constructor compoziție modele IA • Testare de înalt nivel • Arhitectură model • Design hardware model • Reprezentări de date și metadate • Indexări de date
MODELE	**ACTORI/ACȚIONARI**	**DATE**
• Algoritmi • Algoritmi preprocesare date • Algoritmi de instruire • Algoritm selectare subspațiu (caracteristică) • Model • Parametri model • Performanță model • Parametri instruire • Parametri hiper • Modele instruite • Modele reglate	• Proprietar de date • Om de știință de date/dezvoltator IA • Ingineri date • Utilizatori finali • Furnizor/broker de date • Furnizor cloud • Furnizor model • Consumatori servicii/utilizatori model	• Date brute • Set de date etichetate • Set de date publice • Date de instruire • Set de date augmentate • Date de testare • Set de date de validare • Date de evaluare • Set de date preprocesate

(Taxonomia activelor IA)

În încheiere, merită menționat că, datorită complexității IA și sferei extinse a ecosistemului IA, precum și naturii în evoluție a sistemelor și tehnicilor IA, maparea activelor este o sarcină continuă care va avea nevoie de ceva timp pentru a ajunge la un stadiu de maturitate. Acest lucru se datorează unei varietăți de motive/probleme privind natura sistemelor IA (pletor de tehnici și abordări diferite, scenarii diferite de implementare a aplicațiilor, domenii asociate, cum ar fi recunoașterea facială și robotica etc.). O provocare suplimentară implică complexitatea și amploarea lanțului de aprovizionare IA/ML și toate implicațiile pe care acesta le implică pentru peisajul activelor și amenințărilor (4).

Bibliografie

(1) Sfetcu, Nicolae (2022). Introducere în inteligența artificială, Editura MultiMedia Publishing, ISBN 978-606-033-659-4, https://www.telework.ro/ro/e-books/introducere-in-inteligenta-artificiala/

(2) ENISA, *AI Cybersecurity Challenges - Threat Landscape for Artificial Intelligence*, December 2020. Editora: Apostolos Malatras, Georgia Dede – European Union Agency for Cybersecurity. © European Union Agency for Cybersecurity (ENISA), 2020

(3) Pedreschi, D., *Artificial Intelligence (AI): new developments and innovations applied to e-commerce*, Study for the committee on the Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2020. © European Union, 2020

(4) Understanding the Security Implications of the Machine-Learning Supply Chain, https://www.stiftung-nv.de/de/publikation/understanding-security-implications-machine-learning-supply-chain